Clemson University
College of Engineering and Science
Control and Robotics (CRB) Technical Report
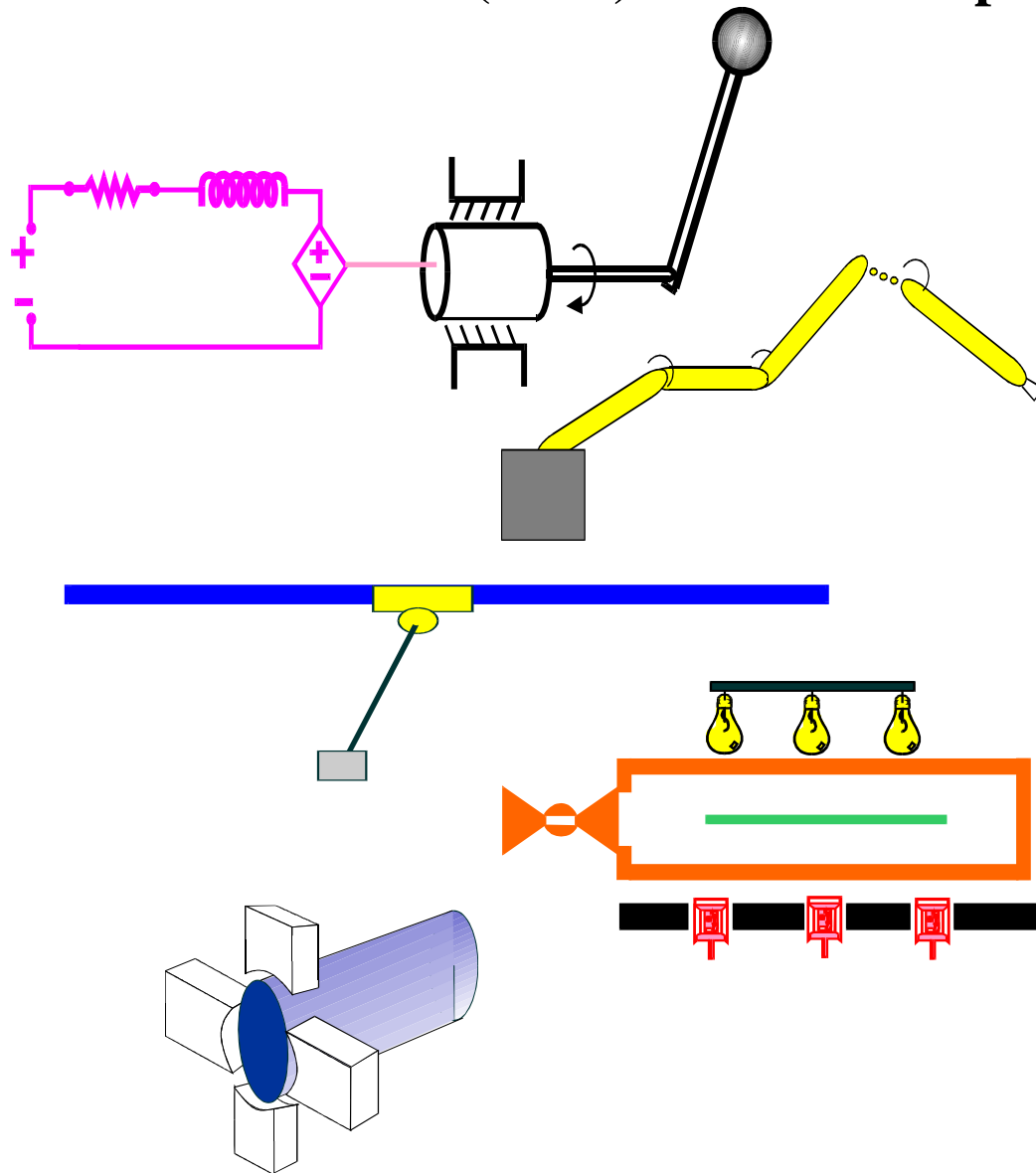
Title: Euclidean Position Estimation of Static Features using a Moving Camera with Known Velocities
Authors: David Braganza, Darren Dawson and Tim Hughes

Report Documentation Page		*Form Approved* *OMB No. 0704-0188*
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.		

1. REPORT DATE **09 MAR 2007**	2. REPORT TYPE	3. DATES COVERED **00-00-2007 to 00-00-2007**
4. TITLE AND SUBTITLE **Euclidean Position Estimation of Static Features using a Moving Camera with Known Velocities**		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Clemson University ,College of Engineering and Science,109 Riggs Hall,Clemson,SC,29631-0901**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**10**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

Euclidean Position Estimation of Static Features using a Moving Camera with Known Velocities

D. Braganza, D. M. Dawson and T. Hughes[*][†]

Abstract

The estimation of 3D Euclidean coordinates of features from 2D images is a problem of significant interest. In this paper we develop a 3D Euclidean position estimation strategy for a static object using a single moving camera whose motion is known. The Euclidean depth estimator which is developed has a very simple mathematical structure and is easy to implement. Numerical simulations and preliminary experimental results using a mobile robot in an indoor environment are presented to illustrate the performance of the algorithm. An extension of this estimation technique for a paracatadioptric system is also presented.

1   Introduction

The use of a camera to estimate the 3D structure of an object from 2D images is known as "Structure from Motion (SFM)" [1, 2, 3], or "Simultaneous Localization and Mapping (SLAM)" (see [4, 5], and references therein). The problem usually involves a camera mounted on a moving platform, such as an unmanned aerial vehicle (UAV) or a mobile robot, which is utilized to map the Euclidean position of static landmarks or visual features in the environment. Recently, SLAM and SFM have been utilized for a number of applications including aerial tracking and surveillance of ground based, stationary or moving objects [6, 7, 8, 9], and terrain mapping systems [10, 11, 12].

Most of the previous results in this area are formulated using linearization based techniques such as the extended Kalman filter [3, 4, 5]. It has been noted [13] that the linearized motion models can cause significant inconsistencies in solutions. There have been a few results [14, 15, 16], which utilized nonlinear system analysis and estimation tools to design nonlinear observers for the problem. In recent work, Chitrakaran et al. [17, 18] proposed nonlinear estimation strategies to identify the Euclidean structure of an object using a monocular calibrated moving camera. The camera motion in this work was modeled based on the homography between two different views captured from the camera, the current frame

and a constant reference frame. The algorithms reported by Chitrakaran et al., require that at least one distance between two features on the object be known for the reconstruction of the 3D Euclidean coordinates. Also, to decompose the homography and obtain the rotation and translation of the camera between the two camera views, the normal vector to the object must be known [18] and in the case of [17], the rotation between the object frame and the camera at the reference position must also be known.

In this work, our objective is to estimate the 3D Euclidean structure of a static object using a single camera mounted on a moving platform whose translation and rotation velocities are measurable. Although the work in [17, 18], was fundamentally more challenging, since the camera velocity was unknown, it did make some assumptions on the structure of the object which it was to identify. There are applications such as video surveillance and mapping using a UAV or a mobile robot where the velocity of the camera mounted on the moving platform is readily available. Thus, the goal of this work is to eliminate the requirements from the previous works [17, 18], that the distance between two feature points be known, and that the normal vector or rotation matrix be known *a priori.* The development in this work is similar to the concepts introduced in [19, 20], where range observer's were developed for feature points on an object undergoing affine motion with known motion parameters. However, the development in our work is based on the kinematics of the moving camera and has a simpler mathematical formulation.

To design the estimator, the equations for motion kinematics are first developed in terms of Euclidean and image-space information based on a single camera view [21]. Then, a nonlinear integral observer [22], is utilized to estimate the velocity of each feature point in the image plane. Once the estimate of the image velocity is known a simple estimator can be developed for the depth variable, and hence, the 3D structure can be estimated. The developed estimator asymptotically identifies the Euclidean depth subject to an observability condition. This condition is similar to the observability condition of [19, 20] and the persistency of excitation condition in [18]. The proposed estimator was implemented using a camera mounted on a mobile robot and our preliminary experimental results show that the estimator converges very quickly and is not computationally complex, and hence, can be used for real-time applications. As an extension

---

[†]The authors are with the Department of Electrical and Computer Engineering, Clemson University, Clemson, SC 29634-0915. `dbragan@ces.clemson.edu`

of this work, we show how the proposed Euclidean estimation strategy can be applied to a paracatadioptric system.

The remainder of this paper is organized as follows, in Section 2, the geometric model which relates Euclidean coordinates of visual features on the stationary object with their corresponding image pixel coordinates is developed based on the perspective projection model. Section 3, describes the motion kinematics between the camera and the object. Section 4, describes the velocity estimator which is used to estimate the pixel coordinate velocity of the visual features, and in Section 5, the Euclidean depth estimator is developed. Finally, numerical simulation and preliminary experimental results using a mobile robot in an indoor environment are presented in Section 6 and Section 7, respectively. In the appendix an extension of this estimation strategy to a Paracatadioptric system is presented.

2   Geometric Model

Figure 1 shows the geometric relationship between a moving perspective camera and features on a static object in its field of view. The geometric model developed in this section is based on a single view of the object from the camera at a time varying position denoted by $\mathcal{I}$. The vector $\bar{m}_i(t) \in \mathbb{R}^3$ denotes the 3D Euclidean position of the ith feature point O_i relative to the camera frame $\mathcal{I}$, and is defined as

$$\bar{m}_i \triangleq \begin{bmatrix} x_i & y_i & z_i \end{bmatrix}^T. \tag{1}$$

The image coordinates of the feature points as captured
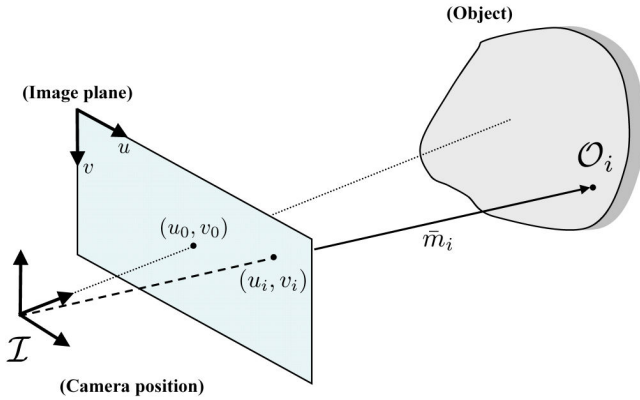

Figure 1: Geometric relationship.

by the moving camera are the normalized Euclidean coordinates of the feature points, denoted by $m_i(t) \in \mathbb{R}^3$ which is defined as

$$m_i \triangleq \frac{1}{z_i} \bar{m}_i. \tag{2}$$

The corresponding projective pixel coordinates of the feature points are denoted by $p_i(t) \in \mathbb{R}^3$ which is defined as follows

$$p_i = \begin{bmatrix} u_i & v_i & 1 \end{bmatrix}^T. \tag{3}$$

The image coordinates of the features and their normalized Euclidean coordinates are related by the pin-hole camera model [21] such that

$$p_i = A m_i \tag{4}$$

where $A \in \mathbb{R}^{3\times 3}$ is a known, constant, and invertible intrinsic camera calibration matrix defined as follows [23]

$$A = \begin{bmatrix} fk_u & fk_u cot(\phi) & u_0 \\ 0 & \frac{fk_v}{sin(\phi)} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{5}$$

where $u_0, v_0 \in \mathbb{R}$ denote the pixel coordinates of the principal point (i.e., the image center that is defined as the frame buffer coordinates of the intersection of the optical axis with the image plane), $k_u, k_v \in \mathbb{R}$ represent camera scaling factors, $\phi \in \mathbb{R}$ is the angle between the camera axes, and $f \in \mathbb{R}$ denotes the camera focal length.

3   Camera Kinematics

The kinematics of the camera frame $\mathcal{I}$ is developed in terms of the image coordinates of the feature points. After taking the time derivative of (4), the following kinematics can be obtained (see [24] for more details)

$$\dot{p}_i = -\frac{1}{z_i} A_{ei} v_c + A_{ei} \left[A^{-1} p_i \right]_\times \omega_c \tag{6}$$

where $v_c, \omega_c \in \mathbb{R}^3$ denote the translational and rotational velocities of the camera relative to the initial position of the camera but expressed in the local frame $\mathcal{I}$, and $A_{ei} \in \mathbb{R}^{3\times 3}$ is a function of the intrinsic camera calibration matrix and the image coordinates of the ith feature points image coordinates, defined as follows

$$A_{ei} = A - \begin{bmatrix} 0 & 0 & u_i \\ 0 & 0 & v_i \\ 0 & 0 & 0 \end{bmatrix}, \tag{7}$$

and $[\zeta]_\times$ denotes the following skew symmetric matrix

$$[\zeta]_\times = \begin{bmatrix} 0 & -\zeta_3 & \zeta_2 \\ \zeta_3 & 0 & -\zeta_1 \\ -\zeta_2 & \zeta_1 & 0 \end{bmatrix} \quad \forall \zeta = \begin{bmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \end{bmatrix}. \tag{8}$$

For the remainder of this development only the first two elements of $\dot{p}_i(t)$ defined in (6) are considered. Thus, the 2D kinematics for the camera can be written as

$$\dot{X}_i = -\frac{1}{z_i} \Pi_i v_c + \Pi_i \left[A^{-1} p_i \right]_\times \omega_c \tag{9}$$

where $X_i(t) \in \mathbb{R}^2$ is expressed as

$$X_i = \begin{bmatrix} u_i & v_i \end{bmatrix}^T \tag{10}$$

and $\Pi_i \in \mathbb{R}^{2\times 3}$ consists of the first two rows of the matrix A_{ei} which was defined in (7), and can be explicitly written as follows

$$\Pi_i = \begin{bmatrix} fk_u & fk_u cot(\phi) & u_0 - u_i \\ 0 & \frac{fk_v}{sin(\phi)} & v_0 - v_i \end{bmatrix}. \quad (11)$$

4 Image Feature Velocity Estimation

The only unknown in the camera kinematic equation (9), is the Euclidean depth $z_i(t)$. To facilitate the development of an estimator for the depth parameter, an estimate of the image velocity signal $\dot{X}_i(t)$ is required. The following continuous estimator [22] can be utilized to estimate the velocity

$$\dot{\hat{X}}_i \triangleq \int_{t_0}^t \left[(K_i + I_2)\tilde{X}_i(\tau) + \Gamma_i \text{sgn}(\tilde{X}_i(\tau)) \right] d\tau + (K_i + I_2)\tilde{X}_i(t) \quad (12)$$

where $\dot{\hat{X}}_i(t) \triangleq \begin{bmatrix} \dot{\hat{u}}_i & \dot{\hat{v}}_i \end{bmatrix}^T \in \mathbb{R}^2$ denotes the estimate of the signal $\dot{X}_i(t)$, $\tilde{X}_i(t) \in \mathbb{R}^2$ is the estimation error defined as follows

$$\tilde{X}_i(t) \triangleq X(t) - \hat{X}(t), \quad (13)$$

$K_i, \Gamma_i \in \mathbb{R}^{2\times 2}$ denote constant positive definite diagonal gain matrices, and $\text{sgn}(\tilde{X}_i)$ denotes the signum function applied to each element of the vector $\tilde{X}_i(t)$. For more details on the development of the above estimator the reader is referred to [22]. To summarize the result, it was shown that the estimator in (12) asymptotically identifies the signal $\dot{X}_i(t)$ (i.e., $\left\| \dot{\tilde{X}}_i(t) \right\|, \left\| \tilde{X}_i(t) \right\| \to 0$ as $t \to \infty$), provided that the jth diagonal element of the gain matrix Γ_i and the jth element of the vectors $\ddot{X}_i(t)$ and $\dddot{X}_i(t)$ satisfies the following condition for all i feature points

$$[\Gamma_i]_j \geq \left| \left[\ddot{X}_i(t) \right]_j \right| + \left| \left[\dddot{X}_i(t) \right]_j \right| \qquad \forall j = 1, 2. \quad (14)$$

Thus the only restriction on the camera motion is a relatively mild assumption of the smoothness and boundedness of the higher order derivatives of the camera velocity.

5 Euclidean Depth Estimation

The objective is to design an estimator for the Euclidean depth, $z_i(t)$. To this end, the kinematic equation (9), can be rewritten in a simplified form as follows

$$\dot{X}_i = -\rho_i\lambda_i + \delta_i \quad (15)$$

where $\lambda_i = \begin{bmatrix} \lambda_{i1} & \lambda_{i2} \end{bmatrix}^T \in \mathbb{R}^2$, $\delta_i = \begin{bmatrix} \delta_{i1} & \delta_{i2} \end{bmatrix}^T \in \mathbb{R}^2$ are measurable signals which are defined as follows

$$\lambda_i = \Pi_i v_c \quad (16)$$
$$\delta_i = \Pi_i \left[A^{-1}p_i \right]_\times \omega_c \quad (17)$$

and $\rho_i(t) = \dfrac{1}{z_i(t)} \in \mathbb{R}$ is the inverse of the Euclidean depth which is unknown and will be estimated.

The individual components of the simplified expression for the camera kinematics in (15), can be written as

$$\dot{\tilde{X}}_{i1} + \dot{\hat{X}}_{i1} = -\rho_i\lambda_{i1} + \delta_{i1} \quad (18)$$
$$\dot{\tilde{X}}_{i2} + \dot{\hat{X}}_{i2} = -\rho_i\lambda_{i2} + \delta_{i2} \quad (19)$$

where $\tilde{X}_i = \begin{bmatrix} \tilde{X}_{i1} & \tilde{X}_{i2} \end{bmatrix}^T$, $\hat{X}_i = \begin{bmatrix} \hat{X}_{i1} & \hat{X}_{i2} \end{bmatrix}^T$, and (13) was utilized. After multiplying (18) by $\lambda_{i1}(t)$ and (19) by $\lambda_{i2}(t)$, and rearranging the resulting equations, the following expressions can be obtained

$$\rho_i\lambda_{i1}^2 = \lambda_{i1}\left(\delta_{i1} - \dot{\hat{X}}_{i1}\right) - \lambda_{i1}\dot{\tilde{X}}_{i1} \quad (20)$$
$$\rho_i\lambda_{i2}^2 = \lambda_{i2}\left(\delta_{i2} - \dot{\hat{X}}_{i2}\right) - \lambda_{i2}\dot{\tilde{X}}_{i2}. \quad (21)$$

After adding (20) and (21), the following expression is obtained

$$\rho_i\left(\lambda_{i1}^2 + \lambda_{i2}^2\right) = \lambda_{i1}\left(\delta_{i1} - \dot{\hat{X}}_{i1}\right) + \lambda_{i2}\left(\delta_{i2} - \dot{\hat{X}}_{i2}\right) \\ -\lambda_{i1}\dot{\tilde{X}}_{i1} - \lambda_{i2}\dot{\tilde{X}}_{i2}. \quad (22)$$

Based on the expression in (22), an estimate for the inverse Euclidean depth can be designed as follows

$$\hat{\rho}_i \triangleq \frac{1}{(\lambda_{i1}^2 + \lambda_{i2}^2)}\left[\lambda_{i1}\left(\delta_{i1} - \dot{\hat{X}}_{i1}\right) + \lambda_{i2}\left(\delta_{i2} - \dot{\hat{X}}_{i2}\right)\right] \quad (23)$$

where $\hat{\rho}_i(t) \in \mathbb{R}$ represents the inverse depth estimate and the inverse depth estimation error $\tilde{\rho}_i(t) \triangleq \rho_i(t) - \hat{\rho}_i(t) \in \mathbb{R}$ is explicitly defined as follows

$$\tilde{\rho}_i = \frac{-1}{(\lambda_{i1}^2 + \lambda_{i2}^2)}\left[\lambda_{i1}\dot{\tilde{X}}_{i1} + \lambda_{i2}\dot{\tilde{X}}_{i2}\right]. \quad (24)$$

Notice that, since the image feature velocity estimator asymptotically converges to the true velocity (i.e., $\dot{\tilde{X}}_{i1}(t), \dot{\tilde{X}}_{i2}(t) \to 0$), the inverse depth estimation error converges to zero, (i.e., $\tilde{\rho}_i(t) \to 0$). Thus, the inverse depth estimate $\hat{\rho}_i(t)$, converges to its true value provided that, $\dot{\hat{X}}(t) \to \dot{X}(t)$ and $\left(\lambda_{i1}^2(t) + \lambda_{i2}^2(t)\right) \neq 0$. From (15), it is evident that, if $\left(\lambda_{i1}^2(t) + \lambda_{i2}^2(t)\right) = 0$, then the inverse depth estimate $\rho_i(t)$ is unobservable. Thus, we can conclude that the inverse depth estimate can be asymptotically identified provided that $\lambda_{i1}^2(t) + \lambda_{i2}^2(t) \neq 0$ and the gain condition in (14) is satisfied.

6 Simulation Results

A simulation study was conducted to evaluate the performance of the proposed estimation algorithm. The simulations were performed using five static feature points

whose Euclidean coordinates were selected as follows

$$
\begin{aligned}
O_1 &= \begin{bmatrix} 0 & 0.2 & 1 \end{bmatrix}^T \\
O_2 &= \begin{bmatrix} -0.1 & 0.2 & 1.25 \end{bmatrix}^T \\
O_3 &= \begin{bmatrix} 0.1 & 0.2 & 1.5 \end{bmatrix}^T \\
O_4 &= \begin{bmatrix} -0.2 & 0.2 & 1.75 \end{bmatrix}^T \\
O_5 &= \begin{bmatrix} 0.2 & 0.2 & 2 \end{bmatrix}^T.
\end{aligned}
\tag{25}
$$

The cameras translational and rotational velocities were chosen as

$$
\begin{aligned}
v_c &= \begin{bmatrix} 0.2\cos(t) & 0.2\sin(t) & 0.1\sin(t) \end{bmatrix}^T \text{ m/s} \\
\omega_c &= \begin{bmatrix} 0 & 0 & 0.1\sin(0.2\pi t) \end{bmatrix}^T \text{ rad/s}.
\end{aligned}
\tag{26}
$$

In addition, a camera calibration matrix for a 640×480 camera was selected as follows

$$
\begin{bmatrix} 810 & 0 & 320 \\ 0 & 820 & 240 \\ 0 & 0 & 1 \end{bmatrix}.
\tag{27}
$$

The estimator gains were chosen to give the best performance both with and without additive noise and were selected as follows

$$
K_i = diag\{20, 20\}, \quad \Gamma_i = diag\{3, 3\}.
\tag{28}
$$

In the simulations four different cases were considered using the above parameters. For case 1, the image points had no noise added to them. In case 2, a small amount of noise (variance 0.001) was added to the image points. For case 3, noise with a variance of 0.0001 was added and image points were passed through a low-pass filter. The low pass filter had a cutoff frequency of 2 Hz. In the final case, the image points were rounded to integers to simulate the discrete output of the feature tracker, these image points were then passed through the low pass filter.

The simulation results for each of the four cases is shown in Table 1. Note that, feature points that are further from the camera generally have a larger error. The highest percent error was 4.3% for case 2 with the feature point at a distance of 2 m from the camera. The depth estimation error for the four cases considered in the simulations are shown in figures 2, 3, 4, 5. Figure 6 shows a comparison of the depth estimation error using the current algorithm and the algorithm from [18] for a single feature point.

7 Experimental Results

In this section, preliminary experimental results using a mobile robot are discussed. A standard off the shelf webcam (Logitech QuickCam) was used to capture images at a resolution of 640×480 pixels. The calibration matrix of the camera was found to be the following

$$
\begin{bmatrix} 726.6 & 0 & 333.3 \\ 0 & 760.8 & 226.2 \\ 0 & 0 & 1 \end{bmatrix}.
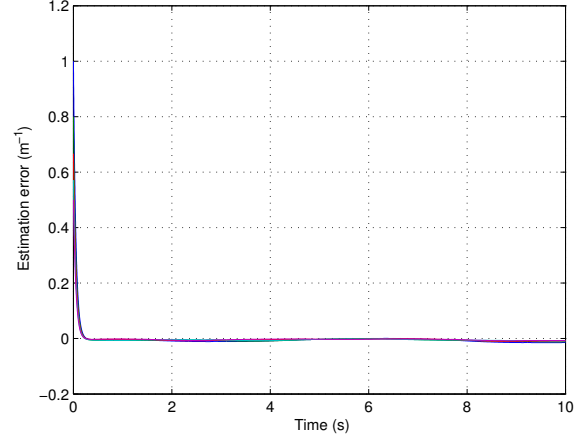\tag{29}
$$

Figure 2: Simulation case 1: Depth estimation error without additive noise.
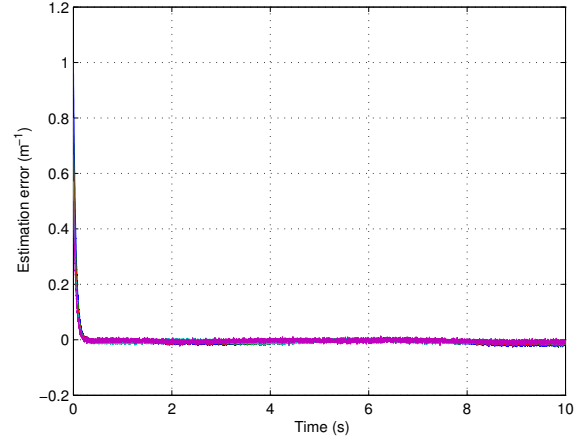

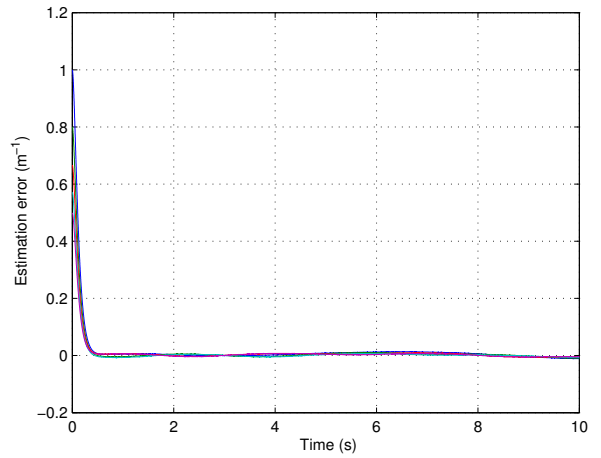Figure 3: Simulation case 2: Depth estimation error with noise of variance 0.001.

Figure 4: Simulation case 3: Depth estimation error with noise of variance 0.0001 and filtering.
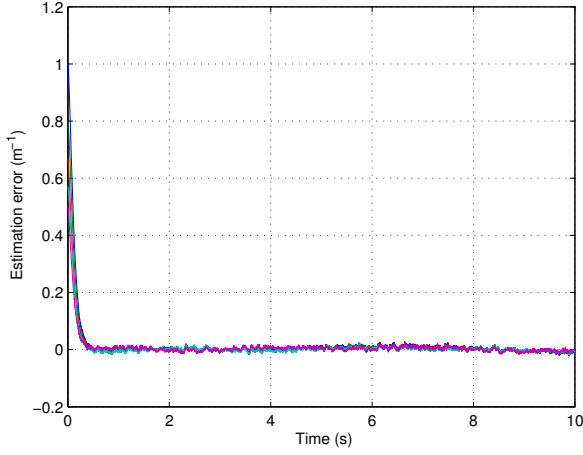
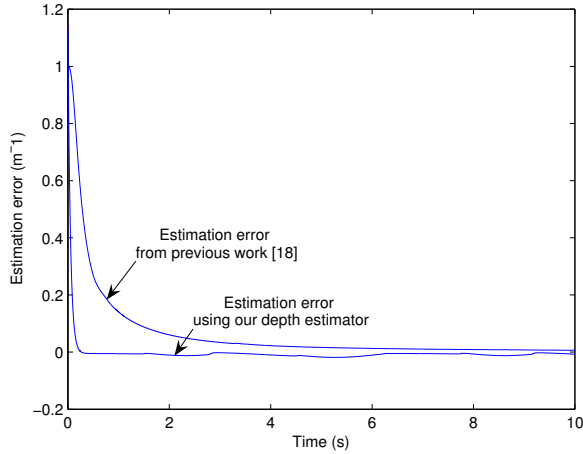Figure 5: Simulation case 4: Depth estimation error with integer rounding.

Figure 6: Simulation comparison of the depth estimation error for a single feature point.

Table 1: Simulation errors in depth estimation for each case

Feature point	Depth (m)	Error Case 1 (cm)	Error Case 2 (cm)	Error Case 3 (cm)	Error Case 4 (cm)
O_1	1.0	1.6	3.0	1.5	1.5
O_2	1.25	2.0	4.1	2.4	2.5
O_3	1.50	2.2	5.3	2.2	2.6
O_4	1.75	2.7	6.9	3.9	4.6
O_5	2.0	3.0	8.5	3.7	4.9

The camera was mounted on an ActivMedia Robotics Pioneer 3 mobile robot as shown in Figure 7. The mobile robot's on-board controller provides translational and rotational velocity information using wheel mounted optical encoders. The test scene consisted of a doll house. Both the mobile robot and the camera were connected to a laptop with an Intel Centrino Duo 2 GHz processor and 1 GB of memory. The laptop was used to set the velocity of the robot, capture images of the scene, and log the video and velocity data for off-line processing. For the preliminary tests, the robot was given a translational velocity of 5 cm/s along the X-axis and no rotational velocity. The average frame rate obtained using the webcam was 14.2 frames per second.
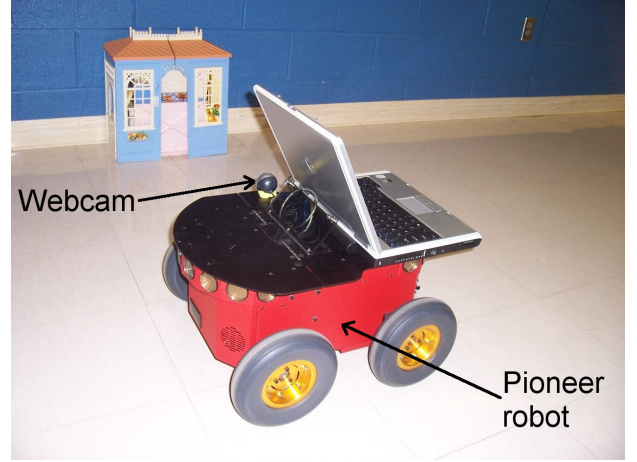

Figure 7: Experimental test setup with camera, mobile robot, and doll-house scene.

Using the implementation of the Lucas-Kanade feature tracking algorithm provided in the OpenCV computer vision library [25], a computer program was written in C++ which enabled the user to select features manually and track those features for the entire image sequence. The program created a text file which contained the feature point pixel coordinates and camera velocity for each frame. In the experiment, 12 features were selected. A sample frame with the tracked feature points is shown in Figure 8.

Figure 8: A frame from the doll-house image sequence showing the tracked feature points.

The depth estimation was calculated off-line using Mathworks Simulink program. A low-pass filter with a cutoff frequency of 1 Hz was applied to the feature points and camera velocities from the text file to smooth the data. The following velocity estimator gains provided good performance

$$K_i = diag\{5, 5\}, \quad \Gamma_i = diag\{1, 1\}. \tag{30}$$

The estimated distance between features is shown in Figure 9. Note that the estimated values stabilize in under 1 second. The estimation error is shown in Figure 10. To illustrate how the image velocity estimator is useful, the image velocity estimator was replaced with a derivative operator. The distance estimation error was seen to be much higher without the image velocity estimator. In fact, the estimation error more than doubled when the velocity estimator was replaced with the simple derivative operation. Figure 11 shows the estimated distance between features, and Figure 12, shows the estimation error for the case with the derivative operation.

Table 2: Experimental error in distance estimation

Object	Actual distance (cm)	Estimated distance (cm)
Length I	10.0	9.49
Length II	23.7	23.4
Length III	40.0	39.6
Length IV	33.7	33.2
Length V	24.5	24.2
Length VI	24.5	24

Figure 10: Experiment: Distance estimation error.

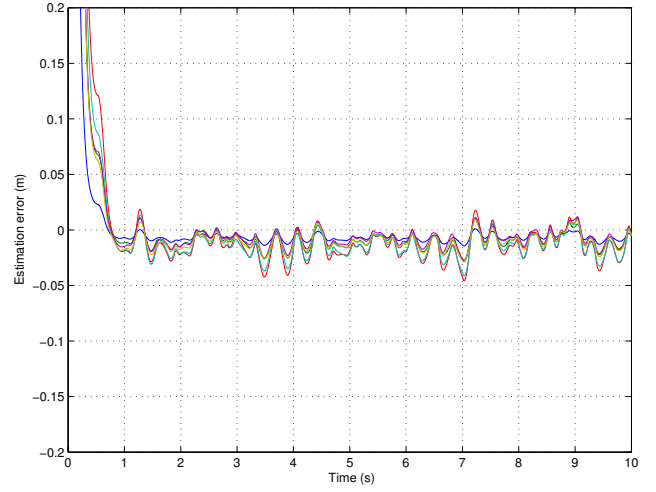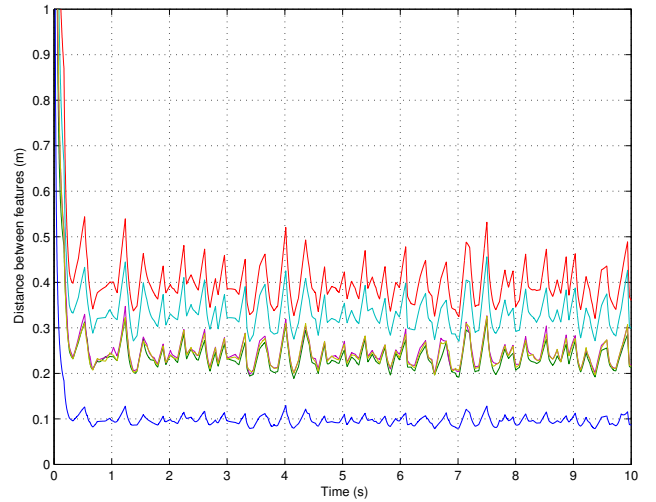Figure 11: Experiment: Estimated distance between features with derivative operator.

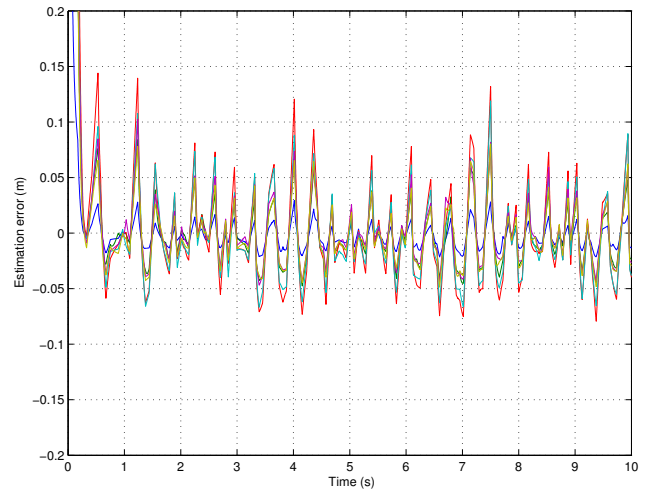Figure 9: Experiment: Estimated distance between features.

Figure 12: Experiment: Distance estimation error with derivative operator.

8   Summary and Future Work

In this paper, we have presented an estimation strategy for 3D Euclidean reconstruction of static features on an object using a single moving camera whose velocities are known. The proposed estimator has a simple mathematical structure and can be easily implemented. Numerical simulations and preliminary experimental results using a mobile robot in an indoor environment were presented. These results demonstrate that the estimation strategy is accurate and converges quickly, in under one second, even with a poor resolution and low frame rate camera. The proposed estimation strategy was also shown to be applicable to a paracatadioptric system. Our future work will consider more complex trajectories for the mobile robot and utilization of a high definition video camera which is expected to provide more accurate results. Further experimental validation using a video camera mounted on a UAV is also being considered. The real-time performance of this algorithm is also currently being evaluated.

APPENDIX

A   Extension for a Paracatadioptric System

A paracatadioptric system is a central catadioptric system with a paraboloid mirror and an orthographic camera [26, 27]. This type of system has a large field of view and thus is suitable for structure from motion. In this section we consider the Euclidean reconstruction of a static object using the 2D images captured from a paracatadioptric system mounted on a moving platform whose velocity is measurable. The development presented is for a single feature point.

The projection of the Euclidean coordinate $\bar{m}(t) = \begin{bmatrix} x & y & z \end{bmatrix}^T \in \mathbb{R}^3$ onto a paraboloid mirror is denoted by the coordinate $\bar{y}(t) = \begin{bmatrix} y_1 & y_2 & y_3 \end{bmatrix}^T \in \mathbb{R}^3$ which is defined as follows [28]

$$\bar{y} \triangleq r\bar{m} \qquad (31)$$

where $r(t) \in \mathbb{R}$ is defined as follows

$$r = \frac{2f}{l} \qquad (32)$$

where $f \in \mathbb{R}$ denotes the constant known focal length of the system and $l(\bar{m}) \in \mathbb{R}$ is the following nonlinear function

$$l = -z + \sqrt{x^2 + y^2 + z^2}. \qquad (33)$$

Note that, since the projection from the mirror onto the camera's sensor is orthographic the signal $\bar{y}(t)$ is directly measurable.

After taking the time derivative of (31), the following expression for the kinematics of $\bar{y}(t)$ can be obtained in a similar manner to [28]

$$\dot{\bar{y}} = \Omega_1 + g \qquad (34)$$

where we have used the following expression to describe the motion of the Euclidean co-ordinate $\bar{m}(t)$ as seen from the moving camera frame [24]

$$\dot{\bar{m}} = -v_c - [\omega_c]_\times \, \bar{m}. \qquad (35)$$

In (35), $v_c, \omega_c \in \mathbb{R}^3$ denote the translational and rotational velocity of the camera frame, and $[\cdot]_\times$ was defined in (8). In (34), $\Omega_1(t) \in \mathbb{R}^3$ consists of measurable and known signals and $g(t) \in \mathbb{R}^3$ contains the unknown Euclidean information which must be estimated. The signals $\Omega_1(t), g(t)$, are defined as follows

$$\Omega_1 = -[\omega_c]_\times \, \bar{y} - \frac{1}{2f} \left(-y_2 \omega_{c1} + y_1 \omega_{c2} \right) \bar{y} \qquad (36)$$

$$g = -r v_c + \left(\frac{v_{c3}}{l} - \frac{y_1 \dot{x} + y_2 \dot{y} + y_3 \dot{z}}{2f(l + x_3)} \right) \bar{y}. \qquad (37)$$

Similar to the development in Section 4, the velocity estimator [22] can be used to develop an estimate of the signal $\bar{y}(t)$ as follows

$$\dot{\hat{\bar{y}}} \triangleq \int_{t_0}^t [(K + I_3) \, \tilde{\bar{y}}(\tau) + \Gamma \mathrm{sgn}(\tilde{\bar{y}}(\tau))] d\tau$$
$$+ (K + I_3) \, \tilde{\bar{y}}(t) \qquad (38)$$

where $K, \Gamma \in \mathbb{R}^{3 \times 3}$ denote constant positive definite diagonal gain matrices, and $\tilde{\bar{y}}(t) \triangleq \bar{y}(t) - \hat{\bar{y}}(t)$ denotes the estimation error for the signal $\bar{y}(t)$. The estimator in (38) asymptotically identifies the signal $\dot{\bar{y}}(t)$ (i.e., $\left\| \dot{\tilde{\bar{y}}}(t) \right\|, \|\tilde{\bar{y}}(t)\| \to 0$ as $t \to \infty$).

Since $\Omega_1(t)$ is measurable, an estimate $\hat{g}(t) = \begin{bmatrix} \hat{g}_1 & \hat{g}_2 & \hat{g}_3 \end{bmatrix}^T \in \mathbb{R}^3$ for the signal $g(t)$ can be defined as follows

$$\hat{g} = -\Omega_1 + \dot{\hat{\bar{y}}} \qquad (39)$$

where $\tilde{g}(t) \triangleq g(t) - \hat{g}(t)$ denotes the estimation error for the signal $g(t)$. Thus, the signal $g(t)$ is asymptotically identified provided that $\dot{\hat{\bar{y}}}(t) \to \dot{\bar{y}}$. From (37), an expression for the unknown Euclidean information contained in $r(t)$ can be developed as follows [28]

$$\hat{r}^2 = \frac{\begin{array}{c} (y_2 \hat{g}_1 - y_1 \hat{g}_2)^2 + (y_3 \hat{g}_1 - y_1 \hat{g}_3)^2 \\ + (y_3 \hat{g}_2 - y_2 \hat{g}_3)^2 \end{array}}{\begin{array}{c} (y_2 v_{c_1} - y_1 v_{c_2})^2 + (y_3 v_{c_1} - y_1 v_{c_3})^2 \\ + (y_3 v_{c_2} - y_2 v_{c_3})^2 \end{array}} \qquad (40)$$

where $\hat{r}(t) \in \mathbb{R}$ denotes an estimate for $r(t)$, and $v_c(t) = \begin{bmatrix} v_{c_1} & v_{c_2} & v_{c_3} \end{bmatrix}^T$. The expression in (40) is obtained by considering the individual elements of (37) and factoring out the second term. Thus the 3D Euclidean information can be reconstructed from (31) provided that the denominator of (40) is non-zero.

References

[1] J. Oliensis, "A critique of structure from motion algorithms," *Computer Vision and Image Understanding*, vol. 80, no. 2, pp. 172–214, 2000.

[2] F. Chaumette, S. Boukir, P. Bouthemy, and D. Juvin, "Structure from controlled motion," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, no. 5, pp. 492–504, May 1996.

[3] A. Chiuso, P. Favaro, H. Jin, and S. Soatto, "Structure from motion causally integrated over time," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, no. 4, pp. 523–535, Apr. 2002.

[4] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: Part i," *IEEE Robot. Automat. Mag.*, vol. 13, no. 3, pp. 99–108, 2006.

[5] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "Monoslam: Real-time single camera slam," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 29, no. 6, June 2007.

[6] T. Fukao, K. Fujitani, and T. Kanade, "An autonomous blimp for a surveillance system," in *Proc. IEEE Int. Conf. Intelligent Robots and Systems*, Las Vegas, NV, 2003, pp. 1820–1825.

[7] T. Kanade, O. Amidi, and Q. Ke, "Real-time and 3d vision for autonomous small and micro air vehicles," in *Proc. IEEE Int. Conf. Decision and Control*, Dec. 2004, pp. 1655–1662.

[8] J. D. Redding, T. W. McLain, R. W. Beard, and C. N. Taylor, "Vision-based target localization from a fixed-wing miniature air vehicle," in *Proc. American Control Conf.*, Minneapolis, MN, 2006, pp. 2862–2867.

[9] V. N. Dobrokhodov, I. I. Kaminer, K. D. Jones, and R. Ghabcheloo, "Vision-based tracking and motion estimation for moving targets using small uavs," in *Proc. American Control Conf.*, Minneapolis, MN, 2006, pp. 1428–1433.

[10] J.-H. Kim and S. Sukkarieh, "Airborne simultaneous localisation and map building," in *Proc. IEEE Int. Conf. Robots and Automation*, Taipei, Taiwan, 2003, pp. 406–411.

[11] I.-K. Jung and S. Lacroix, "High resolution terrain mapping using low altitude aerial stereo imagery," in *Proc. IEEE Int. Conf. Computer Vision*, Nice, France, 2003, pp. 1820–1825.

[12] I. Miyagawa and K. Arakawa, "Motion and shape recovery based on iterative stabilization for modest deviation from planar motion," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 28, no. 7, pp. 1176–1181, July 2006.

[13] S. J. Julier and J. K. Uhlmann, "A counter example to the theory of simultaneous localization and map building," in *Proc. IEEE Int. Conf. Robots and Automation*, Seoul, Korea, 2001, pp. 4238–4243.

[14] M. Jankovic and B. K. Ghosh, "Visually guided ranging from observations of points, lines and curves via an identifier based nonlinear observer," *Systems and Control Letters*, vol. 25, pp. 63–73, 1995.

[15] X. Chen and H. Kano, "State observer for a class of nonlinear systems and its application to machine vision," *IEEE Trans. Automat. Contr.*, vol. 49, no. 11, pp. 2085–2091, Nov. 2004.

[16] X. Hu and T. Ersson, "Active state estimation of nonlinear systems," *Automatica*, vol. 40, p. 2075 2082, 2004.

[17] V. K. Chitrakaran, D. M. Dawson, J. Chen, and H. Kannan, "Velocity and structure estimation of a moving object using a monocular camera," in *Proc. American Control Conf.*, Minneapolis, MN, 2006, pp. 5159–5164.

[18] V. K. Chitrakaran and D. M. Dawson, "A lyapunov-based method for estimation of euclidean position of static features using a single camera," in *Proc. American Control Conf.*, New York, NY, 2007, to appear.

[19] W. E. Dixon, Y. Fang, D. M. Dawson, and T. J. Flynn, "Range identification for perspective vision systems," *IEEE Trans. Automat. Contr.*, vol. 48, no. 12, pp. 2232–2238, 2003.

[20] D. Karagiannis and A. Astolfi, "A new solution to the problem of range identification in perspective vision systems," *IEEE Trans. Automat. Contr.*, vol. 50, no. 12, pp. 2074–2077, 2005.

[21] O. Faugeras, *Three-Dimensional Computer Vision*. Cambridge, Massachusetts: MIT Press, 1993.

[22] V. Chitrakaran, D. M. Dawson, W. E. Dixon, and J. Chen, "Identification of a moving objects velocity with a fixed camera," *Automatica*, vol. 41, no. 3, pp. 553–562, Mar. 2005.

[23] E. Malis and F. Chaumette, "2 1/2 d visual servoing with respect to unknown objects through a new estimation scheme of camera displacement," *Int. Journal of Computer Vision*, vol. 37, no. 1, pp. 79–97, 2000.

[24] J. Chen, D. M. Dawson, W. E. Dixon, and A. Behal, "Adaptive homography-based visual servo tracking for a fixed camera configuration with a camera-in-hand extension," *IEEE Trans. Contr. Syst. Technol.*, vol. 13, no. 5, pp. 814–825, Sept. 2005.

[25] Open source computer vision library. [Online]. Available: http://www.intel.com/technology/computing/opencv/

[26] S. Baker and S. K. Nayar, "A theory of single-viewpoint catadioptric image formation," *Int. Journal of Computer Vision*, vol. 35, no. 2, pp. 175–196, 1999.

[27] C. Geyer and K. Daniilidis, "Paracatadioptric camera calibration," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, no. 5, pp. 687–695, May 2002.

[28] D. Aiken, S. Gupta, G. Hu, and W. E. Dixon, "Lyapunov-based range identification for a paracatadioptric system," in *Proc. IEEE Int. Conf. Decision and Control*, San Diego, CA, 2006, pp. 3879–3884.